STOCK MARKET PREDICTION USING MACHINE LEARNING

### By

**Name - ARNAB KARMAKAR**

**Reg No - 19BAI1090**

**ABSTRACT**

The financial exchange is one of the basic areas of a nation's economy. It permits financial backers to contribute and acquire profits from their venture. Foreseeing the securities exchange is an extremely difficult errand and has drawn in serious interest from specialists from many fields like measurements, man-made consciousness, financial matters, and money. An exact forecast of the financial exchange decreases speculation risk on the lookout. Various methodologies have been utilized to foresee the securities exchange. The exhibitions of Machine learning (ML) models are commonly better than those of measurable and econometric models. The brain network is one of the smart information mining procedures that has been involved by analysts in different regions for the beyond 10 years. Expectation and examination of financial exchange information play had a significant impact in the present economy. The different calculations utilized for gauging can be sorted into direct (AR, MA, ARIMA, ARMA) and non-straight models (ARCH, GARCH, Neural Network). Yet, this likewise intends that there's all's a ton of information to track down designs in. Thus, monetary investigators, specialists, and information researchers continue to investigate examination procedures to distinguish securities exchange patterns. This led to the idea of algorithmic exchanging, which utilizations computerized, pre-modified exchanging techniques to execute orders. I'll utilize both conventional quantitative money system and AI calculations to foresee stock developments. We'll go through the accompanying points:

Stock examination: major versus specialized investigation

Stock costs as time-series information and related ideas

Anticipating stock costs with Moving Average strategies

Prologue to LSTMs

Foreseeing stock costs with a LSTM model

Last considerations on new philosophies, like ES

## PROBLEM STATEMENT

This part examines the difficulties that were experienced in finishing this undertaking. Every issue is momentarily depicted and examined exhaustively in resulting areas of the review. The most importantly challenge was removing and dealing with the authentic information of insider exchanging. The verifiable information ought to be adequately sufficiently enormous to prepare the model and back test the model for its exactness. Dealing with the information included separating the information to just incorporate important information, and amalgamating the different exchanges organization wise in the month-to-month time period to make the information fit for preparing. This was accomplished utilizing Excel capabilities like channel and turn table. We likewise needed to remove the authentic qualities of organizations that were pertinent at the hour of purchasing. It was hard to scrap information through various locales particularly stock-related destinations as there is a lot of futile data accessible, picking the one you really want is truly troublesome. Finding the locales which contain the data you want is another difficult assignment. When we had the sifted advertiser purchasing information and properties of the organization, we expected to settle on the design of the model. The construction that should have been made depended on monetary hypothesis. We needed to settle on questions like what time periods ought to be incorporated, should there be a benchmark rate, and if indeed, what ought to be the benchmark rate for various time spans?

**INTRODUCTION**

1.1 INTRODUCTION

This section aims to provide a concise overview of the project, including the project's goals,

challenges encountered, and report structure.

1.2 PROJECT OVERVIEW

Investment is a crucial part of finance. People/Companies invest in various assets to generate

returns. Stocks are the most famous asset class for investment purposes. Returns are generated

through capital appreciation or dividend payments in the stock market. However, stocks also

carry a risk component, the price of the stock can go down resulting in losses. Thus, it is

important to correctly select the stock that you will invest in, in order to generate a positive

return.

In order to generate a positive return, one must be able to predict the price security should hold

in the future and invest/choose not to invest in it based on its price. Analysts generally use

Fundamental and/or Technical analysis to come to an intrinsic value for the stock.

AI is largely being adopted in the finance space. It is playing a major role in many important

decisions making, models, investments, etc. This project aims to build an AI-based model that

will help select stocks for investment purposes. We used seven machine learning and four deep

learning algorithms. Since, the number of input data will gradually increase with respect to time

will increase so neural network algorithms will perform better as accuracy of the neural network

model increase with the increase in the input data.

The model should be able to predict stocks that are expected to appreciate in the near future. For

correct prediction, we will be giving out promoter buying data combined with other attributes of

the company. The rationale being a promoter is the best judge of the company and will not buy a

substantial amount of stock if he doesn't expect it to do well in the future. The other attributes of

the company are fed to increase the accuracy of the model.

[12]A corporate promoter is a firm or person who does the preliminary work related to the

formation of a company, including its promotion, incorporation, and flotation, and solicits people

to invest money in the company, usually when it is being formed. In most cases, the promoter is the founder and operator of the company. Therefore, a promoter is expected to have the best

knowledge in terms of both company and the market.

1.3 PROJECT GOALS

The finalized project should be able to achieve the following purposes:

● Check if it is possible to create a successful portfolio with the promoter buying sheet.

● Create models that give buy/not buy decisions.

● Check how the model works in different time frames.

● Check if the portfolio gives alpha during the holding period.

● Check the impact of changing the return benchmark from zero on accuracy and portfolio

return.

1.4 PROJECT CHALLENGES

This section discusses the challenges that were encountered in completing this project.

Each problem is briefly described and discussed in detail in subsequent sections of the

study.

The first and foremost challenge was extracting and managing the historical data of insider

trading. The historical data should be sufficiently large enough to train the model and backtest

the model for its accuracy. Managing the data included filtering the data to only include relevant

data, and amalgamating the various transactions company-wise in the monthly time frame to

make the data fit for training. This was achieved using Excel functions like filter and pivot table.

We also had to extract the historical attributes of companies that were relevant at the time of

buying. It was difficult to scrap data through different sites especially stock-related sites as there

is too much useless information available, picking the one you need is really difficult. Finding

the sites which contain the information you need is another challenging task.

Once we had the filtered promoter buying data and attributes of the company, we needed to

decide on the structure of the model. The structure that needed to be created was based on

financial theory. We had to decide on questions like what timeframes should be included, Should

there be a benchmark rate, and If yes, what should be the benchmark rate for different time

periods.

**LITERATURE SURVEY**

2.1 RESEARCH OVERVIEW

This segment intends to provide information about the undertaken research related to the project.

The chapter includes research from various sources about varied topics required to understand

the project. This includes knowledge about stock prediction using insider data. This chapter will

also talk about the technology used and the data prepared in building the project.

2.2 DATASET RESEARCH

Construction of models of machine learning and deep learning needs a heavy amount of data or

datasets to build on efficiently. Before the beginning of any task of the project, the dataset must

be thoroughly researched. The first of the tasks was to identify the datasets to further build data

on. For the project, the main datasets that have been chosen are the last 6-year promoter buying

dataset and then collected information regarding the stocks which will help in better prediction

2.3 EXISTING TECHNOLOGY SOLUTIONS

In the following section a comprehensive review of the STOCK PREDICTION from the

literature is done.

[1]Alan M. Safer has proposed the application of neural networks to predict abnormal stock

returns using insider trading data. The research covered 343 companies over the period of 4

years. The research resulted in a few findings which are i) increasing the time of the future

forecast the accuracy increases, ii) to increase the accuracy the time-period of back aggregated

data should be increased. The paper mostly focused on a particular industry type thus focusing

on a single industry type helped in better training of the algorithm. As well as the paper

narrowed its assessment to small and mid-size companies rather than large ones.

[2]Kolasani S.V. and Assaf R. have proposed a system to predict stock movement with help of

sentiment analysis using a Twitter feed with the help of a neural network. They attempted to

forecast future stock market movement in the United States by assessing the emotion of

market-related Twitter messages. Using the SVM algorithm, they gathered relevant tweets and

calculated their average sentiment value. They then created a training set containing those tweets

and the matching Apple Inc or DJIA closing stock index differential between today and

tomorrow and tested it on comparable stock-related tweets on a separate timeline to see how well

they could forecast the stock index. To forecast the stock index, they employed a Multilayer

Perceptron Neural Network model and a Boosted Regression Tree model. It was discovered that

tweets play an important part in the forecast of stock movement and that the Neural Network

outperforms the Boosted Regression Tree in predicting too high and too low discrepancies in the

stock index.

[3]Kohli P.P.S., Zargar S., Arora S., and Gupta P. tried to predict the behavior of the Bombay

Stock Exchange (BSE) using machine learning algorithms. They used elements including

commodity prices (crude oil, gold, silver), the foreign exchange rate (FEX), and market history

to anticipate the behavior of the Bombay Stock Exchange as input attributes for multiple

machine learning models (BSE). Because the correlation factor is the strongest, the data show

that BSE is most dependent on the gold rate. Furthermore, the silver rate has the lowest

correlation factor, indicating that BSE has the least reliance on it. The AdaBoost method has the

greatest accuracy of 76.79 percent for 70 percent of training data and 75 percent for untrained

data out of all the machine learning algorithms examined.

[4]Pang X., Zhou Y., Wang P., Lin W., and Chang V. tried to improve stock market forecasts, and

for that, an innovative neural network technique was developed. Data was obtained live from the

stock market for real-time and off-line analysis, as well as the results of visualizations and

analyses, to demonstrate the Internet of Multimedia of Things for stock analysis. They used a

deep long short-term memory neural network (LSTM) with an embedded layer and a long

short-term memory neural network with an automatic encoder to anticipate the stock market. In

the trials, the deep LSTM with an embedded layer performs better. The accuracy of the two

models for the Shanghai A-shares composite index is 57.2 percent and 56.9 percent, respectively.

[5]Urmita Sharma, and Sumedha Seniaray proposed a deep learning-based model to make

prediction more reliable and simpler. The LSTM technique, which is an advanced type of

Recurrent Neural Network, was the focus of the article. They were able to predict the stock's

closing price for the next 10 days after doing the experiment. They assessed the model's

accuracy on 80 percent of the data that was already available and on 20 percent of the data that

was not yet available. They also used an ADAM optimizer to improve it.